Title
Automated Multiscale Approach To Predict Self-Diffusion from a Potential Energy Field.

Journal
Journal of chemical theory and computation, 15(4)

ISSN
1549-9618

Authors
Mace, Amber
Barthel, Senja
Smit, Berend

Publication Date
2019-04-01

DOI
10.1021/acs.jctc.8b01255

Peer reviewed

JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2019, 15, 2127−2141

pubs.acs.org/JCTC

Automated Multiscale Approach To Predict Self-Diffusion from a Potential Energy Field

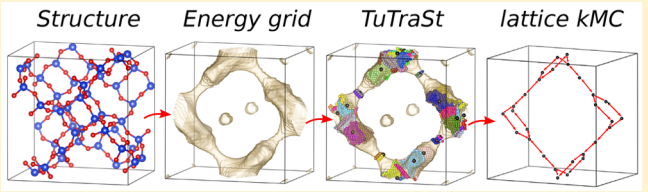
Amber Mace,[‡,¶,†] Senja Barthel,[‡,†] and Berend Smit*[,‡]

[‡]Laboratory of Molecular Simulation (LSMO), Institut des Sciences et Ingénierie Chimiques, Valais, Ecole Polytechnique Fédérale de Lausanne (EPFL), Rue de l'Industrie 17, CH-1951 Sion, Switzerland

[¶]Department of Materials and Environmental Chemistry, Stockholm University, SE-106 91 Stockholm, Sweden

Supporting Information

ABSTRACT: For large-scale screening studies there is a need to estimate the diffusion of gas molecules in nanoporous materials more efficiently than (brute force) molecular dynamics. In particular for systems with low diffusion coefficients molecular dynamics can be prohibitively expensive. An alternative is to compute the hopping rates between adsorption sites using transition state theory. For large-scale screening this requires the automatic detection of the transition states between the adsorption sites along the different diffusion paths. Here an algorithm is presented that analyzes energy grids for the moving particles. It detects the energies at which diffusion paths are formed, together with their directions. This allows for easy identification of nondiffusive systems. For diffusive systems, it partitions the grid coordinates assigned to energy basins and transitions states, permitting a transition state theory based analysis of the diffusion. We test our method on CH_4 diffusion in zeolites, using a standard kinetic Monte Carlo simulation based on the output of our grid analysis. We find that it is accurate, fast, and rigorous without limitations to the geometries of the diffusion tunnels or transition states.



1. INTRODUCTION

Modeling diffusion in solid-state materials is important for many applications. One such is the use of nanoporous materials as membranes. Here the ratio of the diffusion coefficients of the gases one would like to separate in these materials is an important factor which, together with the adsorption, determines the selectivity of a membrane.[1−10] Another timely application is identifying superconductive solid-state electrolyte materials for Li-ion batteries. Here, the transition from using flammable, nonaqueous liquid organic electrolytes to solid-state electrolytes is anticipated to open up for extensive improvements of the power density, time stability, and safety of such a battery.[11−13]

During the past decade, high-throughput screening studies have emerged to find and improve the understanding of structure−activity relationships for different classes of materials and to accelerate the process of discovering novel materials for urgently needed, e.g. green, energy applications.[14,15] Theoretical screening of a wide range of materials before synthesis will save huge amounts of time and expenses in this process, significantly increasing our prospects of finding an optimal material for the task. This development has, in turn, entailed the development of crystal structure prediction where databases of millions of hypothetical materials to be screened for different properties are arising[16−21] in addition to those of already synthesized structures.[22−27]

To date, numerous high-throughput screening studies have been carried out on porous materials databases consisting of different nanoporous materials such as zeolites, metal−organic frameworks, covalent-organic frameworks, zeolitic imidazolate, etc. for thermodynamic aspects of CH_4 storage,[18,28−36] CO_2 capture,[28,37−40] and Kr/Xe separation.[41−44] Likewise for solid-state electrolyte materials, extensive screening studies have been carried out focusing on static energetic equilibrium properties such as bandgaps.[20] However, similar studies for dynamic properties like diffusion coefficients are scarce,[3−5] despite being key parameters not only for predicting selectivity or ionic conductivity in a material but also for determining whether the diffusion of gases in porous materials is so slow that diffusion limitation makes thermodynamic analyses of limited use.

A major challenge in screening the diffusion coefficients of guest molecules in solids is the sheer quantity of guest-structure combinations that can be realized. With a nearly limitless compositional space yet to be explored a brute-force screening with a state-of-the-art molecular dynamics (MD) approach would merely scratch the surface. We must therefore employ new strategies to intelligently design our screening procedures to exclude nondiffusive materials and short-list those to be subject for more detailed theoretical studies and finally to be synthesized and tested experimentally.

We are proposing a novel screening method to estimate the diffusive properties of particles in solid-state materials with a multiscale modeling approach. The procedure is based on a geometric isosurface analysis of the discretized three-dimen-

Received: December 13, 2018
Published: February 27, 2019

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

sional scalar field quantifying the potential energy experienced by a migrating particle within a rigid lattice structure. The potential energy can be translated into the free energy of diffusion at infinite dilution. This simplification entails that loading effects or particle−particle interactions are not taken into account. Such a potential energy field can easily be computed using a preparametrized force field or at the density functional theory level of theory.[45] The geometric analysis predicts the activation energy and directions of diffusion while coarse-graining the energy landscape. This coarse-graining quantifies the energy data into a lattice model with particle site-to-site hopping rates. The lattice model acts as input to a kinetic Monte Carlo (kMC) algorithm[46] for computing the diffusion coefficients for even complicated transition state landscapes which can be cumbersome to compute analytically. The procedure we propose does not require any structure specific parametrization.

Transition state theory (TST) based lattice models including kMC have been used extensively to describe gas diffusion in nanopores[9,47−50] and also, to a lesser extent, Li^+ diffusion in solid-state materials.[51] However, these studies use information from MD simulations to predict the pore-to-pore diffusion rates, hence the kMC-algorithm is used solely as a way to accelerate the scale of the simulations and can thus not be used to predict diffusive behavior *a priori* performing a more computationally costly MD simulation. In addition, when sampling transition states of rate determining processes for diffusion rates slower than the MD simulation time scale, the MD is dependent on sampling bias techniques, such as umbrella sampling,[52] constrained MD,[53] metadynamics,[54] or nudged elastic band.[55] Applying these techniques can often be tedious, time-consuming, and impossible to apply in a high-throughput manor.

The idea of analyzing potential energy landscapes for diffusion to derive a lattice model to which TST was applied was presented by Kim et al.[3] who estimated diffusion coefficients for CO_2, N_2, and CH_4 adsorbates in zeolites for gas-separation applications. Similar to our approach, they filled up the energy levels up to a cutoff energy of 15 k_BT to find points that potentially belong to diffusion paths. Kim et al. determined one-dimensional diffusion paths, which involved the approximation that the transition states are perpendicular to the unit cell vectors and restricting to orthogonal cells. In this work we show that these restrictions can be removed by presenting a different approach to identify the diffusion pathways, which allows us to identify transition states in diffusive systems of any shapes and allowing all possible unit cells.

Our TuTraSt algorithm is rigorous in finding all types of channels and TS surfaces allowing us to predict the diffusion with a multiscale modeling approach. The procedure we propose does not require any structure specific parametrization and can be fully automated, and the computational cost is only a fraction of that of an MD simulation.To our knowledge, there is no other such algorithm available to the scientific community.

The anticipated major outcome of making this algorithm available to the computational chemistry community is the facilitation of reliable high-throughput screening studies of guest particle diffusion in solid-state materials. Our goal is to make possible the prediction of millions of diffusion coefficients at a manageable computational cost where >90% are within the experimental accuracy of 1 order of magnitude relative to those computed by MD. We apply the TuTraSt algorithm to the study of CH_4 diffusion in 113 zeolites to prove that this goal is, in fact, realistic and to test the limits. We find that the TuTraSt algorithm reaches the set goals where 96% of the 339 diffusion coefficients were predicted within the accuracy limit and with a speedup of >5000 relative to the MD where the directional diffusion coefficients for a structure can be computed in <1 h on a single CPU.

Finally, the TuTraST algorithm is general in essence as a method of geometric analysis of scalar fields and is not restricted to particle diffusion in solid-state materials. An example is the analysis of bonds based on ideas of the quantum theory of atoms in molecules: A scalar field function of the electron density in a molecule was analyzed and the characterization of chemical bonds facilitated by the TuTraSt algorithm that allows for identification of the precise position and function value at which basins, corresponding to high electron densities, merge.

2. THEORETICAL ASPECTS

The diffusion of a guest particle in a solid-state crystalline host material is determined by a complex interplay between particle−host, particle−particle, and host−host interactions. To study diffusion computationally, the most common approach is to perform MD simulations. This can be done using preparameterized force fields computing the interatomic forces at each integration step or computing the forces "on-the-fly" using density functional theory (DFT), i.e. *ab initio* MD. From the trajectories of N guest particles i, the mean-square displacements (MSD) can be computed (eq 1). Provided that the simulations are over a sufficiently long time-scale to reach the diffusive regime, the self-diffusion coefficients D_s can be computed according to eq 2 for a dimensions. By computing D_s at different temperatures T, we can extrapolate the diffusional activation energy (E_a) and temperature independent pre-exponential D_0 with an Arrhenius plot according to eq 3.

$$\mathrm{MSD} = \langle(\vec{r}(t) - \vec{r}(t_0))^2\rangle = \frac{1}{N}\sum_{i=1}^{N}(\vec{r}_i(t) - \vec{r}_i(t_0))^2 \tag{1}$$

$$D_s = \frac{1}{2a}\lim_{t\to\infty}\frac{d}{dt}\langle(\vec{r}(t) - \vec{r}(t_0))^2\rangle \tag{2}$$

$$D_s = D_0 e^{-E_a/k_BT} \rightarrow \ln D_s = \ln D_0 - E_a/k_BT \tag{3}$$

To compute the activation energy thus requires several separate MD simulations per structure on the nanosecond time scale for diffusion involving activated processes. For limited numbers of structures and using preparameterized force fields, this can be feasible. However, when looking at several thousands of structures, this computational cost will quickly become impractical while diffusivity slower than 10^{-8} cm^2/s and 10^{-4} cm^2/s will become prohibitively expensive for classical and *ab initio* MD, respectively. Due to this computational cost, screening studies on diffusion properties using MD are limited.[4,5] However, from a scientific point of view, the diffusion of a mobile species within a crystalline material can be considered a site-to-site hopping motion, on a rigid lattice of sites. Instead of following the diffusion process in "real-time" such as in an MD simulation, we will energetically consider the mobile particles moving within a three-dimensional potential-energy landscape. In such a case

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

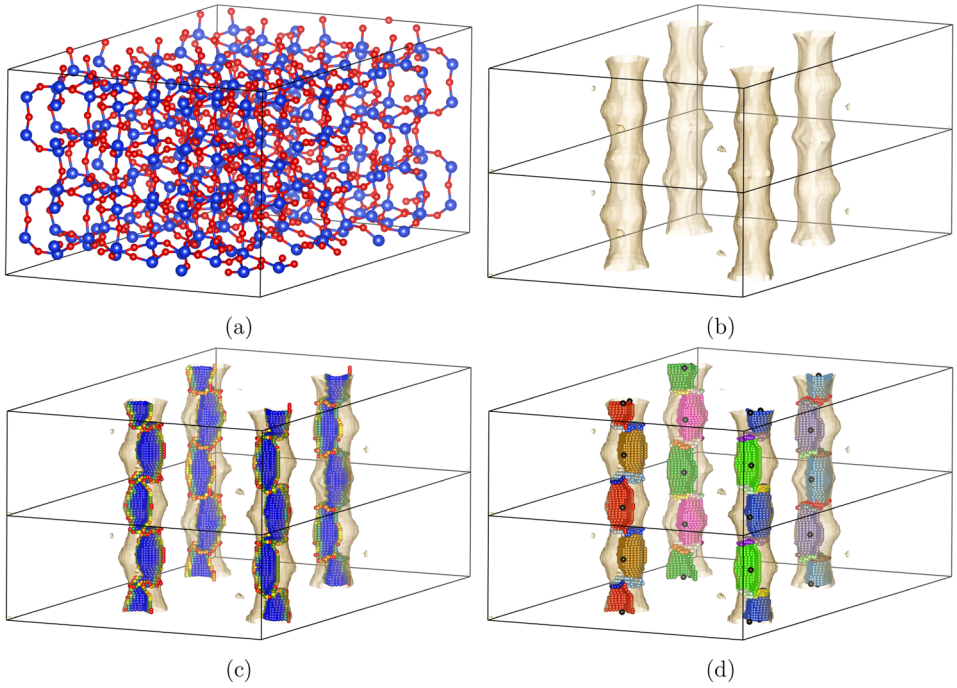


Figure 1. (a) A 2 × 1 × 1 supercell of the IZA structure framework of PSI (blue = Si and red = O atoms). (b) The isosurface of the potential energy at 60 kJ/mol, which defines the accessible pore volume, shows that PSI has a one-dimensional tunnel system accessible to CH_4. (c) The TSs of the one-dimensional CH_4 diffusion are colored by energy levels (blue = low to red = high). (d) The independent TS surfaces dividing basin pairs are colored differently; the black points show the positions that represent the lattice sites used in the kMC simulation.

that energy minima, "basins", can be sufficiently separated energetically by energy barrier dividing surfaces, transition states (TS), the hopping rate between them can be predicted by TST.[52] The time spent in each basin before jumping to the next is "coarse-grained", and with kMC we will model the jump in only one simulation step, independently of the actual time the process is predicted to take. The spatial and temporal propagation of the particle is recorded at each step, producing a trajectory from where the MSD and D_s can be computed according to eqs 1 and 2, respectively, just like in an MD simulation.

Provided that the basins and TS surfaces in the potential energy landscape are known, the Bennet-Chandler approach[52,56,57] allows us to compute the transition rates from basin A to basin B according to eqs 4 and 5.

$$k_{A\to B} = \kappa \times \sqrt{\frac{k_B T}{2\pi m}} \times P(q^*) \tag{4}$$

$$P(q^*) = \frac{\int_{TS} e^{-E(q^*)/k_B T} dq^*}{\int_{\text{basin } A} e^{-E(q)/k_B T} dq} \tag{5}$$

In eq 4 the transmission coefficient κ corresponds to the probability of the particle ending up in basin B when at the top of the barrier. In the case that the identified TS is true, then $\kappa = 1/2$, $\sqrt{T/2\pi m}$ is the averaged velocity at the top of the barrier assuming that the particle follows the Boltzmann distribution at temperature T, and k_B is the Boltzmann constant. The TS surface coordinates are denoted by q^*, and the coordinates in the basin A are denoted by q. $P(q^*)$ is the probability of finding the particle at the top of the barrier, i.e. at a coordinate q^* within the TS surface dividing basins A and B. $P(q^*)$ is computed according to eq 5 where the numerator and denominator are the integrals of Boltzmann sums of the energies E over the coordinates q^* and q, respectively.

Provided that we have a three-dimensional grid quantifying the potential energy felt by the mobile species at each point in space within the host structures, the scientific challenge is to develop the methodology to accurately partition the grid points into basin volumes and transition state surfaces to construct our lattice model. One can then use analytical theory or kinetic Monte Carlo to compute a self-diffusion coefficient from the hopping rates.

The here presented identification of diffusion pathways and the assignment of energies were done differently from the work by Kim et al.:[3] whereas our method identifies basins and TSs for systems independently from the dimension and shape or alignment of tunnels with respect to the cell, Kim et al. determined one-dimensional diffusion paths only that intersect each layer perpendicular to the cell vectors of a unit cell (one dimension at a time) exactly once. The energies along the path were computed as the Boltzmann sums of the points in the cross sections of the path in each layer, and consequently the TSs are approximated as perpendicular to the cell vectors. Kim et al. restricted their analysis to orthogonal cells. The approach furthermore fails to detect paths correctly that are more complicated, such as running back and forward, or diagonal through the cell. The diffusion coefficient is calculated by combining the diffusions along all one-dimensional diffusion paths. Whereas this approach can be used to analyze diffusion paths in well-ordered zeolites with orthogonal cells, the outlined restrictions make it impossible to use it in more complicated systems.

To illustrate some of the materials in which the method of Kim et al. fails to give an accurate description, we have looked at the diffusion in the zeolites PSI, IFY, and AEI, where the three letter abbreviation refers to the International Zeolite

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

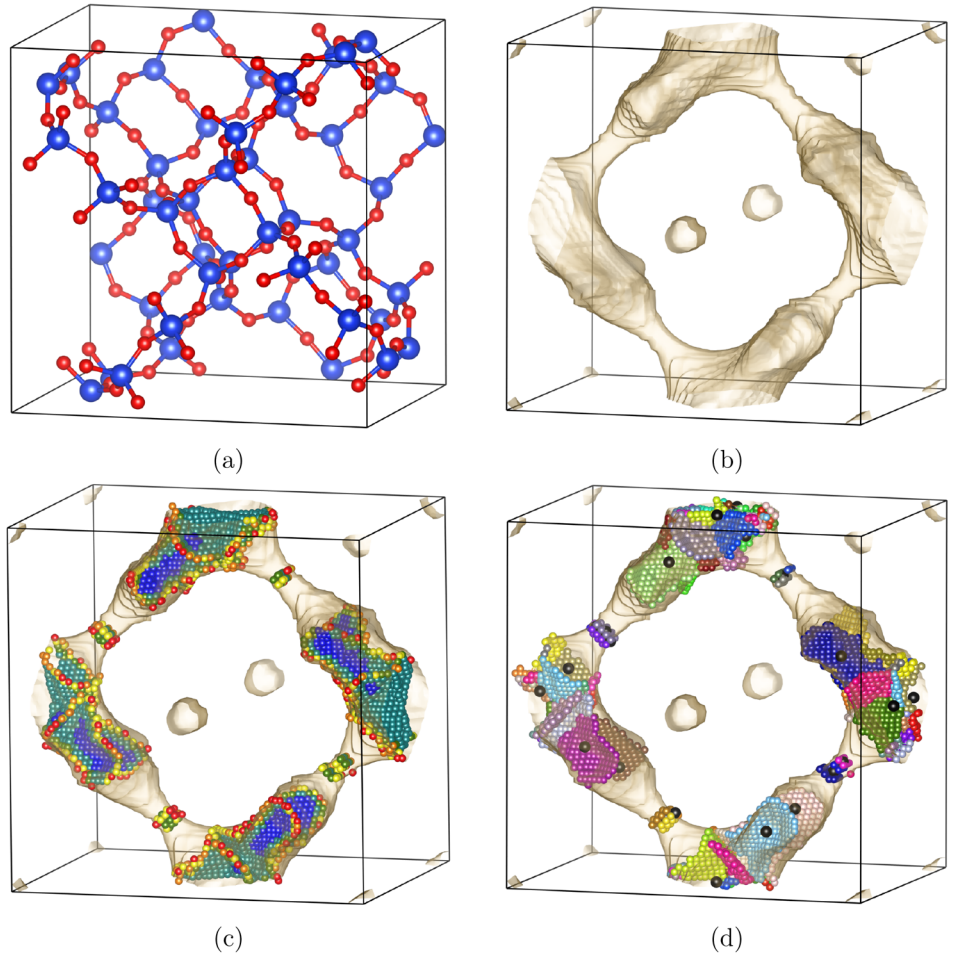


Figure 2. (a) One unit cell of the IZA structure framework of IFY (blue = Si and red = O atoms). (b) The isosurface of the potential energy at 60 kJ/mol shows that IFY has a two-dimensional tunnel system accessible to CH_4. The isolated spherical cavities are inaccessible from the outside and therefore do not play a role in the diffusion. (c) The TSs of the two-dimensional CH_4 diffusion are colored by energy levels (blue = low to red = high). (d) The independent TS surfaces dividing basin pairs are colored differently; the black points show the positions that represent the lattice sites used in the kMC simulation.

Association (IZA) structures[22] which are shown in Figures 1a, 2a, and 3a. The examples clearly show transition states that are not aligned parallel to the cell sides. All three examples furthermore illustrate that the surfaces between basins do not even need to be orthogonal to the center paths of the diffusion system (tunnel).

Figure 1b shows the diffusion system (or tunnel) of the zeolite PSI shown in Figure 1a. The diffusion system of PSI consists of simple one-dimensional channels running parallel to one cell vector that are formed of cavities which are connected along surfaces both orthogonal to the tunnel direction as well as parallel to it, connecting basins sidewise. The partitioning in sidewise connected basins cannot be detected by the method of Kim et al. Figure 1c shows the resulting transition states colored by the energies. In Figure 1d each transition state is colored differently. The coordinates that represent the lattice sites are indicated by the black dots. They are calculated as the center of the basin coordinates, considering the Boltzmann weighting. Figure 2a shows the zeolite IFY which has a two-dimensional or tunnel, and Figure 2b shows how diffusion paths in this zeolite run diagonal in two dimensions inside the unit cell. Again, from Figures 2c and 2d it can clearly be seen that none of the TS surfaces are parallel to the cell sides. Furthermore, several basins together form the junctions of the diffusion system, with boundary surfaces not separating the tunnels orthogonally. A similar example for framework AEI with a three-dimensional tunnel system is given in Figures 3a-3d. At the crossings where diffusion paths that extend in different directions come together, the basins merge in complicated surfaces, defining TSs that are not parallel to the unit cell or orthogonal to a simple path.

We overcome the limits by developing a rigorous algorithm for tunnel detection, and instead of combining one-dimensional diffusions, we detect basins and the TSs as their merging points only depend on the energy values, building a lattice model based on this information. This approach allows us to consider the complete system at once without any restrictions on the choice of the unit cell or the complexity of the diffusion system.

We have thus successfully developed an algorithm that searches for tunnels and transition states (TuTraSt) that iteratively grows minimum energy basins one isosurface layer at a time with a connected component search. It identifies transition state grid points at the position where basins merge. In addition, the TuTraSt algorithm identifies connected components and finds energy levels at which conduction tunnels are formed for each direction, i.e. the diffusional activation energy.

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

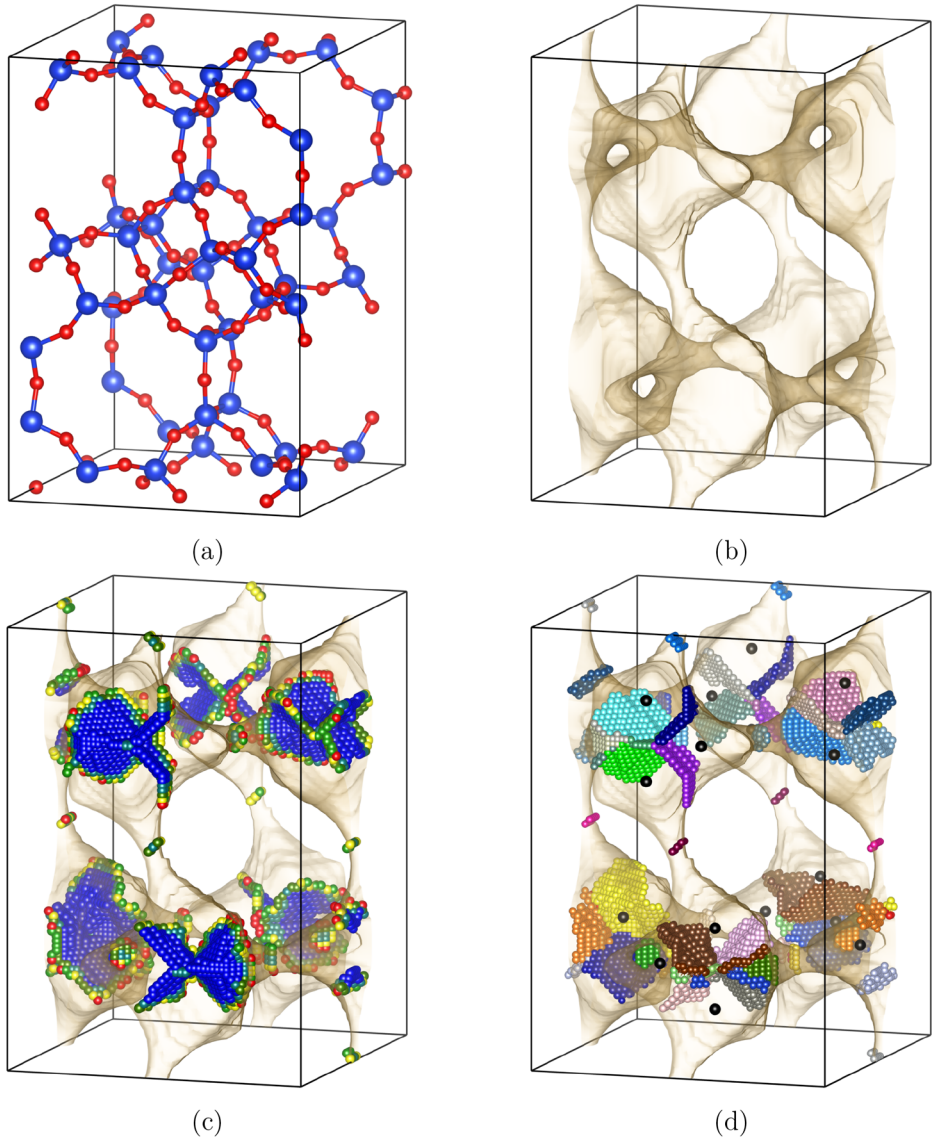


Figure 3. (a) One unit cell of the IZA structure framework of AEI (blue = Si and red = O atoms). (b) The isosurface of the potential energy at 60 kJ/mol shows that AEI has a three-dimensional tunnel system accessible to CH_4. (c) The TSs of the three-dimensional CH_4 diffusion are colored by energy levels (blue = low to red = high). (d) The independent TS surfaces dividing basin pairs are colored differently; the black points show the positions that represent the lattice sites used in the kMC simulation.

3. METHOD

The presented TuTraSt algorithm consists of the following parts: Using an energy grid as input only, it detects tunnel systems together with their dimensions and breakthrough values, identifies basins and transition states (TSs), and from these determines the process rates and coordinates of the basins that can be subsequently used to perform a transition state theory (TST) based analysis to compute the diffusion coefficients, here implemented by a lattice based kMC approach.

The idea of the algorithm is to follow the appearance, growing, and merging of clusters while stepping through the discretized levels of the (potential) energy grid, from the minimal energy up to a chosen maximal value (Figure 4). Clusters will represent basins, and the points on the surfaces between them correspond to TSs.

Detecting Tunnels in 3d Grids. We initialize our calculations by computing the energy of our guest molecule with the atoms of the host lattice. A tunnel is a connected component of the diffusive system, consisting of energy values larger than a given value, that runs through the entire material. A tunnel is called one-dimensional if it extends infinitely into one direction and two-dimensional (three-dimensional) if it contains two (three) paths that extend infinitely into two (three) linearly independent directions.

All grid points with values smaller than or equal to the level are found, and each cluster found by a neighbor search (i.e., each connected component) is analyzed for tunnels (P1−P4 in Figure 4). For that, it is checked whether the cluster connects to a point with one of its periodic images, which can be seen by comparing the boundary crossing vectors of points. The boundary crossing vectors contain the information on which a copy of the unit cell point was first found, i.e. it is counting how often a boundary is crossed with taking directions into account. The location of the reached periodic image defines the direction of the tunnel. The direction is obtained from the

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

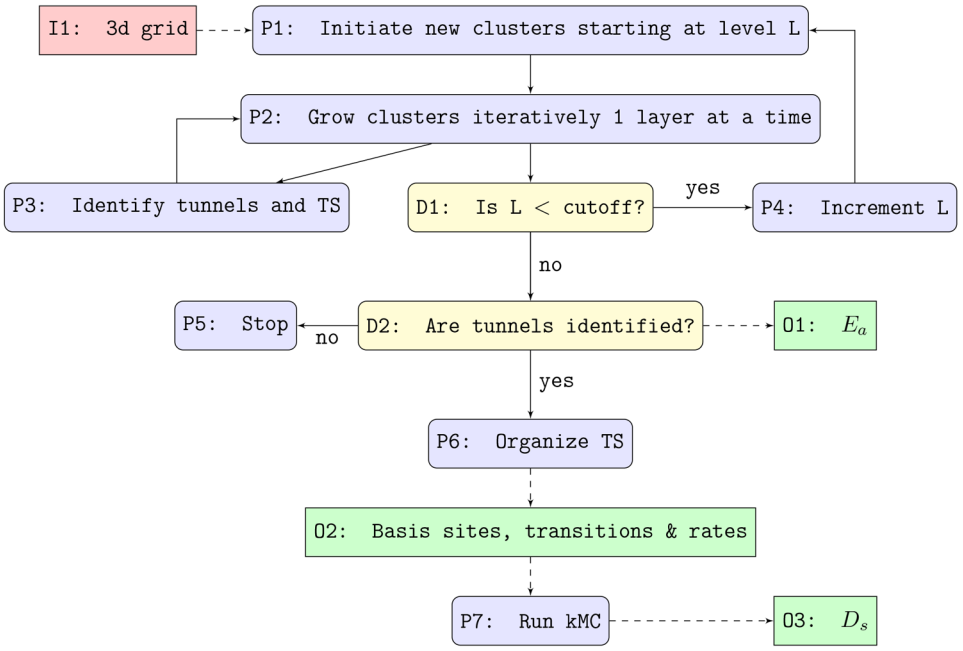


Figure 4. Schematic outline of the TuTraSt algorithm where the boxes are colored by red = input (I1), green = output (O1−O3), blue = processes (P1−P7), yellow = decisions (D1−D2).

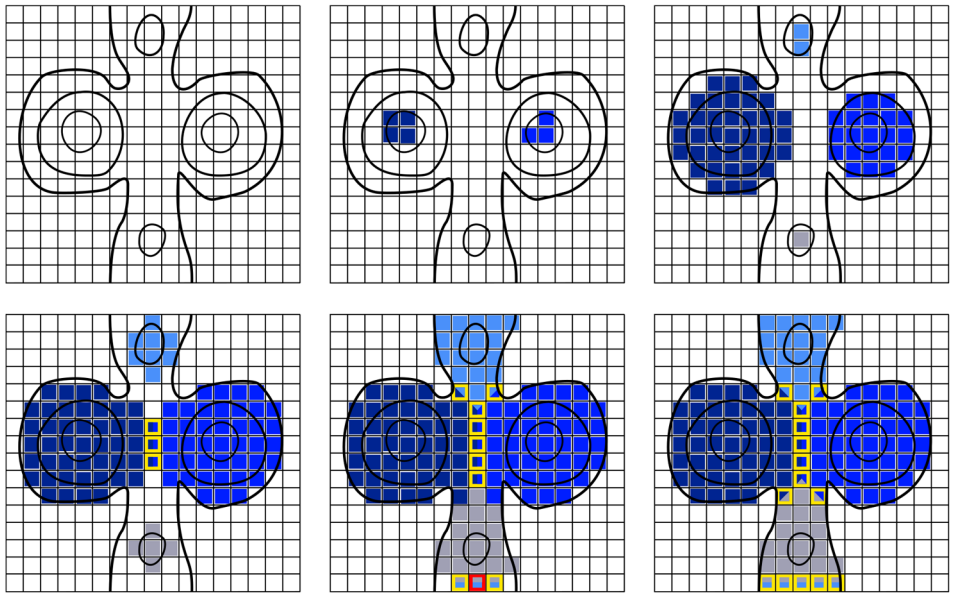


Figure 5. Growth of basins with tunnel and TS identification. The clusters are grown in the order: left, right, top, bottom. A tunnel is detected in the middle figure of the bottom line (red). Points belonging to TSs are indicated with yellow boundaries.

difference of their boundary crossing vectors. To determine tunnels correctly, it is important to update boundary crossing vectors in the case that the point in which two different clusters C_1 and C_2 merge connects points with different boundary crossing vectors b_1 and b_2: In this case the vector $b_1 - b_2$ is added to the boundary vectors of all points in C_2. The number of linearly independent crossing vectors between periodic images in one cluster defines its dimensionality, i.e. the number of directions a tunnel system extends to. The values at which new directions of tunnels per cluster are formed are stored as breakthrough values. These breakthrough values estimate the diffusional activation energy E_a (O1 in Figure 4). The neighbor search considers grid points with common faces only and does not take grid points that only have a corner in common (diagonal checks) into account, since this could lead to errors in the detection of the surfaces between clusters, which are found as described in the following.

Finding Transition States and Basins. A more detailed analyses produces estimates of the basins and transition states for a chosen maximum energy. To obtain the basins and transition states, the TuTraSt algorithm detects merging points that define boundary surfaces between the growing clusters while stepping up the energy levels of the grid (P4 in Figure 4). For that, it is crucial to grow the clusters iteratively in each single level. Otherwise TSs would be assigned incorrectly; in the worst case a complete shell around one cluster would be

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

falsely identified as TSs. Figure 5 shows the iterative growth of four basins (left, right, top, bottom basin) in each level (isolines). TSs are identified as merging points between basins (yellow). The last figure (bottom-right) shows that six TSs are identified. One tunnel is detected with the inclusion of the red point in the middle image of the bottom line in Figure 6. The

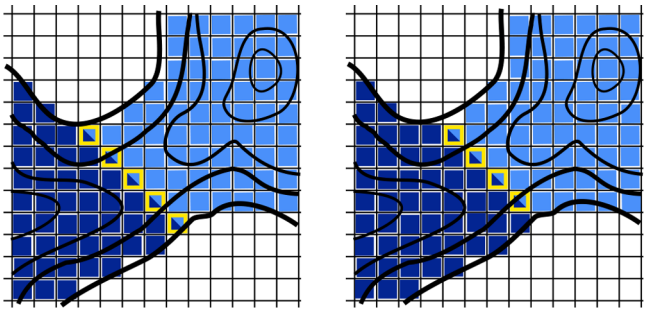


Figure 6. Illustration of the dependency of the position of the TSs on the order in which the basins are grown. No periodic boundary conditions are applied in this figure.

points assigned to TSs in principle depend on the order in which the clusters are grown, but this effect is negligible since the assignment never differs by more than one step in the grid. The boundary surfaces between clusters are assumed to be representative for the transition states if the energy differences between the first merging point and the minima of the two neighboring basins are greater than a given value. Otherwise the barrier is neglected as too small, and the two clusters are merged. Finally, all points belonging to boundary surfaces between the same pair of basins are clustered with a neighbor search. Each of the resulting components gives one transition.

Assignment of Transition Rates and Computing Diffusion Coefficients. To obtain a lattice model that serves as input for a TST based computation of the diffusion, coordinates for the basins and probabilities of the transitions need to be computed (P6 in Figure 4). First, clusters that belong to a tunnel system are identified since only these contribute to diffusion. The positions of the basins are given as their centers of mass. This approximation will only lead to problems if the center of mass lies outside the basin and in another basin - this can only happen in the very unlikely case that one basin is enclosing another one. All connected TS surfaces between neighboring basins are identified, each defining a transition. The direction of a transition is obtained as the difference of the boundary crossing vectors of two energy minima, one in each of the neighboring basins, where the boundary crossing vectors are calculated along a path connecting the two minima that passes through no transition states other than the one of interest. It is important to calculate the directions for each connected component of the boundary surface between a pair of basins, since the TSs in tunnels that contain only one or two basins correspond to distinct transitions. The rate of the transition between two neighboring basins is obtained by taking the Boltzmann weighted sum of all energy values of the points on a TS, as well as for all points in each of the neighboring basins according to the Bennet-Chandler eqs 4 and 5 (O2 in Figure 4).

For simple systems with known noncomplex systems of transitions, the application of TST using eqs 4 and 5 to obtain the diffusion coefficients is analytically straightforward. However, for an automated screening study the complexity of the transition state landscapes was unknown, and it is practical and powerful to use a kMC algorithm which propagates the particle through the lattice stochastically based on the probability for each transition to occur. For our diffusion application we have thus integrated a *rejection-free lattice* kMC simulation[46] (P7 in Figure 4) in the TuTraSt algorithm to automatically compute the directional diffusion coefficients from the lattice model output information (O3 in Figure 4).

4. CASE STUDY: CH_4 DIFFUSION IN ZEOLITES

To test TuTraSt, we apply our algorithm to study CH_4 diffusion in all-silica zeolites. The characteristic one-, two-, or three-dimensional formation of diffusion tunnels in zeolites is periodically separated by pore windows which act as diffusional "bottlenecks" in addition to the potential energy wells at adsorption sites. Passing these energetic barriers can thus be considered rare events in relation to the diffusion within the pore or well, and the pore-to-pore/well-to-well diffusion can be considered a hopping motion to which we can assign a hopping rate making zeolites ideal systems to validate our prediction of diffusion coefficients in each of the 3 dimensions.

From the IZA database,[22] we select all-silica zeolites which are identified by Zeo++[58] to be porous to CH_4 that will constitute our validation set. Porosity is determined for the rigid frameworks since all-silica zeolites are known to be relatively rigid. Flexibility has shown to have the largest effect when the tunnel system contains narrow windows close to the kinetic diameter of the guest molecule showing enhancements of diffusion coefficients of up to 39% which corresponds to less than 0.1 order of magnitude.[59] In our screening these very narrow pores have not been included.

We load 1 CH_4 per unit cell into these 113 structures and model the dynamics of the CH_4 in a rigid framework with MD. The MD simulations are run for 100 ns with a 1 fs time step at 500 K in the NVT ensemble using the Nose-Hoover thermostat with the LAMMPS software.[60,61] The UFF force field parameters[62] are used to describe the nonbonded Lennard-Jones interactions with framework molecules, while the united atom TraPPE force field[63] is used to describe the CH_4 molecule which is modeled effectively as a single particle. The Lennard-Jones parameters between framework atoms and CH_4 are combined using the Lorentz−Berthelot mixing rules. In this model CH_4 does not carry charges, and hence the Coulomb interactions do not need to be computed. From the trajectories where the CH_4 coordinates are printed every 100 fs, the mean-square displacement (MSD) is computed for time windows $(t - t_0)$ up to 10% of the total simulation time and for all possible values of t_0 according to eq 1. This is done for each of the three unit cell basis directions (a, b, c). From the MSD curves the diffusion coefficient D_s of the respective direction is computed according to eq 2 from the slope of a 3 ps region directly following t at which the root mean-square displacement (RMSD) exceeds the length of the respective cell parameter $(\lambda_{a,b,c})$ to confirm that we have reached the time scale at which the CH_4 is diffusing through the unit cell. We further confirm that the diffusive regime has been reached by checking that the slope of MSD as a function of t on the log−log scale is ∼1 and that the D_s is converged within this region by computing D_s for smaller time windows of 1 ps. For directions where the RMSD does not exceed the cell parameter, D_s is set to 0, and we conclude that CH_4 does not diffuse on the time scale of the MD simulation. From the

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

113 structures we produce a total of 339 directional MSD curves. Of these 199 were found to have reached the diffusive regime from which we computed diffusion coefficients (D_s^{MD}).

The three-dimensional potential energy grids for each structure in our validation set are computed with two different voxel sizes 0.1 × 0.1 × 0.1 Å^3 and 0.2 × 0.2 × 0.2 Å^3. This is done by placing a CH_4 in the center of each voxel and summing the Lennard-Jones interaction energy to each atom in the framework using the same force field as for the MD. These grids are then used as input for the TuTraSt algorithm, partitioning the energetic data from the grid into data read by the kMC algorithm which outputs the 339 directional diffusion coefficients (D_s^{TTS}) which we compare with the respective D_s^{MD}. We compute and compare the D_s^{TTS} for the two sets of grids for a range of energy step sizes (0.1, 0.2, 0.5, and 1.0 kJ/mol) and energy cutoff values (40, 60, and 80 kJ/mol) to test the limits of these energetic parameters. In our calculations the time limit for each calculation is set 150× the median compute time per structure of data set 5 (DS5), *vide infra*. In our case, this corresponds to ∼70 h running on a single CPU of an Intel Broadwell processor running at 2.6 GHz. Carrying out TuTraSt analysis on structures requiring compute times beyond this limit is not desirable for screening purposes since it approaches the MD time scale.

The energy step (E_{step}) is the parameter that defines the lower limit of energy barriers which confine the space of a basin and thus the level of detail in which the TS landscape is described by the lattice model. If the energy difference between the lowest energy value within a basin cluster and that within a TS surface separating the basin from a second connected basin is below E_{step}, we consider that these two basins are merged into one. If the energy step is too big, rate determining transitions can be missed. This is particularly significant for fast diffusing materials where the barriers can be missed completely, merging the diffusion tunnel into one single basin, resulting in a failure of the analysis. Setting the energy level parameter too small can on the other hand result in an overidentification of processes which cannot be considered "rare events" with a sufficient separation of the diffusional time scale in relation to movements within the basin. In such a case that processes are included which are not rate determining and are significantly faster than these, the kMC simulation will spend most of its time sampling the fast processes instead of the rate determining processes.

The energy cutoff (E_{cutoff}) defines the maximum energy of grid points which are included in the energetic partitioning. The number of levels (nL) that the algorithm is iterated over is thus $nL = E_{cutoff}/E_{step}$, and the TuTraSt algorithm stops once all grid sites with $E_{grid} \leq E_{cutoff}$ have been analyzed. E_{cutoff} should be set to a value that corresponds to an energy which in relation to the given temperature is the limit where the probability of a particle visiting that state is negligible and the grid point can be considered inaccessible. Setting E_{cutoff} too high results in unnecessarily long computation times and an excess of nonrate-determining processes with negligible probabilities to be executed. This could result in difficulties interpreting the TS surface if required. A too small E_{cutoff} would result in a failure to identify significant diffusion channels and nonconverged processes rates. In the study of Kim et al.[3] the equivalent of E_{cutoff} was set to 15 k_BT which at the simulation temperature of 500 K of the current study is equivalent to 62.0 kJ/mol.

5. RESULTS AND DISCUSSION

In Figure 7 we compare our TST computed diffusion coefficients D_s^{TTS} with the MD results D_s^{MD}. This comparison

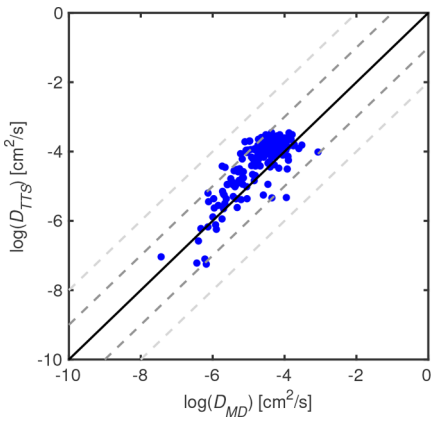


Figure 7. Directional diffusion coefficients computed with the TuTraSt algorithm with optimized energetic parameters (g_{res} = 0.2 Å, E_{cutoff} = 60 kJ/mol, and E_{step} = 0.1 kJ/mol) on the *y*-axis relative to the corresponding diffusion coefficients computed with MD on the *x*-axis on a log−log scale for DS5. The solid black line represents a perfect correspondence between the data sets, while the light gray and dark gray dashed lines guide the limits for deviation of 1 and 2 orders of magnitude, respectively.

shows that for most materials our predictions are within 1 order of magnitude of the exact MD results. Such accuracy is sufficient for most applications.

It is instructive to investigate how our results depend on the different parameters that we can use to tune the accuracy of our calculations: The D_s^{TTS}'s in Figure 8 are computed for energy grid resolutions g_{res} = 0.1, and Figure 9 shows them for g_{res} = 0.2. For the finer g_{res} = 0.1 grid, no significant loss of accuracy is observed when decreasing the detail of our lattice model by increasing E_{step} from 0.1 to 0.2 kJ/mol. However, when increasing E_{step} further to 0.5 and 1.0 kJ/mol, an increasing number of outliers appear, where $D_s^{TuTraSt}$ is underestimated relative to D_s^{MD}. When varying the E_{cutoff} values, no significant change in correspondence between kMC and MD data is observed. For the coarser 0.2 Å grid, outliers arise when increasing the E_{step} from 0.1 to 0.2 kJ/mol, while the correspondence is similarly stable throughout the E_{cutoff} values.

Since the energy step size shows instabilities at 0.2 kJ/mol, we restrict the further analysis of the results to E_{step} = 0.1 kJ/mol. For these 6 data sets (DS1-DS6) of varying E_{cutoff} = 40, 60, and 80 kJ/mol and g_{res} = 0.1 and 0.2 Å, we proceed with a detailed comparison of accuracy, stability, and speed. To facilitate this comparison we present a number of metrics. nUF (unfinished) defines the number of TuTraSt structure calculations ×3 that did not finish within the time limit. nD_0 is the number of directions accurately predicted to be nondiffusive by both MD and TuTraSt as well as those that were predicted diffusive by TuTraSt but not MD, however where the D_S^{TTS} is lower than or within 1 order of magnitude of the estimated lower limit of D_s^{MD} that can be probed within the MD simulation time and is thus within the desired accuracy limit. For the diffusive data for which we have both D_s^{TTS} and D_s^{MD}, nD_{accept} tells us the number of D_s^{TTS} that are within our accuracy goal of 1 order of magnitude, while $nD_{outlier}$ tells us the

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

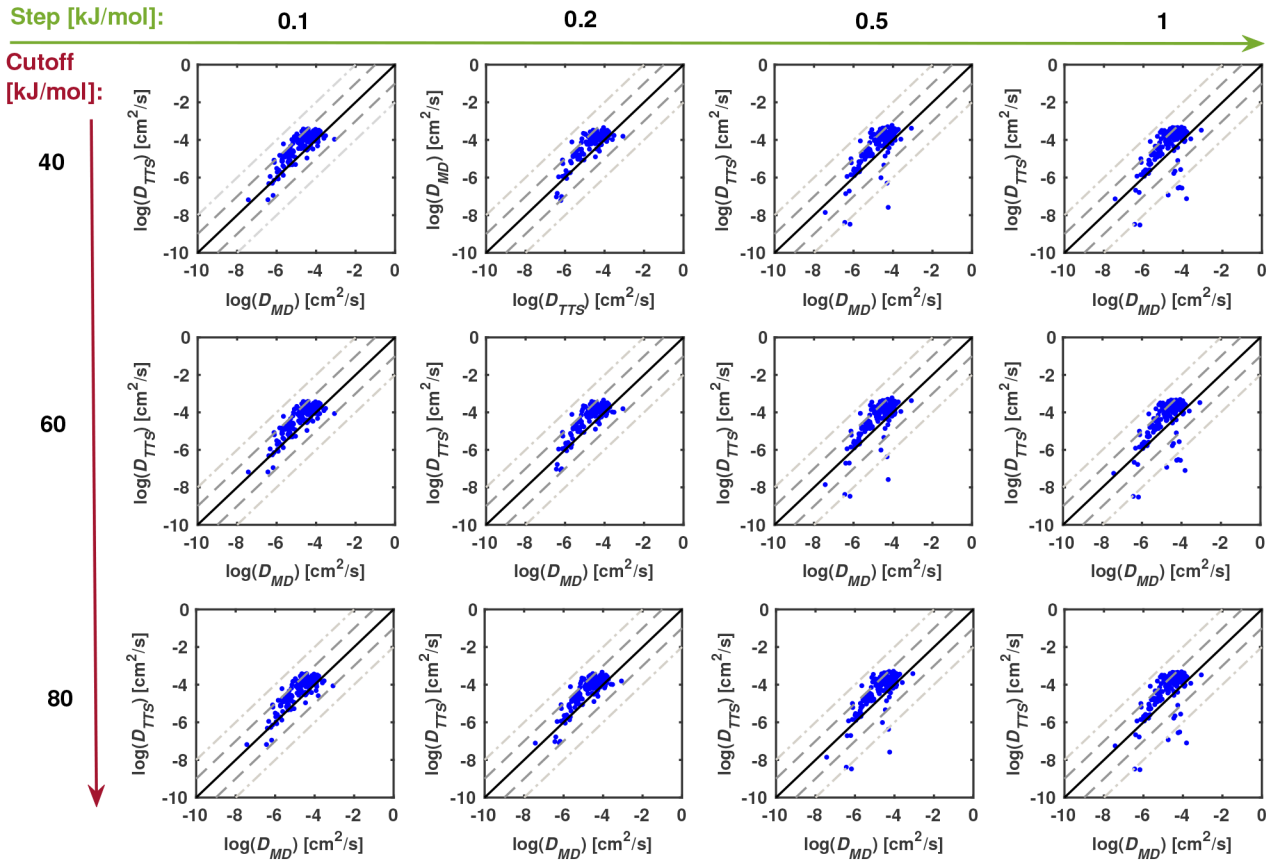


Figure 8. Comparing the D_{TTS} and D_{MD} for all IZA zeolites, using grids with grid size 0.1 Å. The plots differ by the choice of the values of E_{cutoff} and E_{step}.

number of D_s^{TTS} that are outside this limit. To identify systematic deviations we compute the bias $\epsilon = \sum_n \log(D_s^{TTS}/D_s^{MD})/nD_s$ along with the variance $\sigma = (\log(D_s^{kMC}/D_s^{MD}) - \epsilon)^2/nD_s$ of the data. The Spearman's rank correlation coefficient ($\rho_{Spearman}$) is a metric comparing the ranking of diffusion coefficients obtained from the MD and TuTraSt. Identical rankings result in $\rho_{Spearman} = 1$. We define the percentage ($\%_{accept}$) of the total 339 diffusion directions (nD_{tot}) for which the diffusion has been predicted within our defined accuracy limits i.e. $\%_{accept} = 100 \times (nD_0 + nD_{accept})/nD_{tot}$. *SU* notates the mean speedup for the TuTraSt computation time relative to the corresponding MD computation time per structure. Table 1 presents these metrics for the 6 data sets.

In Figures 10 and 11 the data for DS2 and DS5 are presented. Here the spread of the D_s^{TTS} data relative to the D_s^{MD} validation data is shown in the left and middle panels, while the right panel shows the spread of the speedup of the TuTraSt calculations relative to the MD, per structure.

From the results presented in Table 1, we can conclude that the speedup when using the TuTraSt algorithm to compute the diffusion coefficient relative to MD is drastic. Further by increasing the grid resolution from g_{res} = 0.1 to 0.2 Å the number of grid points is decreased by a factor of 8. This speeds up, in particular, the neighborhood search which is the computational bottleneck in several processes of the TuTraSt algorithm, namely P2 (see Figure 4) for the cluster growth and P6 for grouping connected TS points into separate transitions and finding transition directions. Comparing the speedup of DS5 relative to that of DS2 shows that using the coarser grid results in a further significant speedup of 13× allowing the TuTraSt calculations to finish within our set time limit while a loss of accuracy is not noted. The use of a coarser grid also decreases the memory requirements for the TuTraSt calculations, which can facilitate the analysis of structures with larger unit cells.

Using E_{cutoff} = 40 kJ/mol decreases the computational costs relative to E_{cutoff} = 60 kJ/mol; however, it shows to be at the limit of the energies sampled during an MD simulation within a nanosecond time scale leading to a few diffusion channels not being identified at energies with breakthrough energies >40 kJ/mol, while diffusion was shown in these directions by the MD. Looking at the higher cutoff value E_{cutoff} = 80 kJ/mol there is no accuracy improvement, while the computational time increases and fewer of the TuTraSt calculations finished within the time limit, in turn, decreasing the total $\%_{accept}$. We thus confirm that a $E_{cuttoff}$ of 60 kJ/mol (~15 k_BT) is appropriate when comparing with MD as suggested by Kim et al. However, if a study is performed where one wishes to detect slow diffusion beyond the MD time scale, use of a higher E_{cutoff} will be necessary. From the results we also conclude that an E_{step} value of 0.1 kJ/mol is necessary to withhold the stability of the calculations, while the loss in speed is insignificant relative to E_{step} = 0.2 kJ/mol (<10%). For each new class of materials it is however advisible to run a representative subset against MD to calibrate the energy parameters.

Following this discussion we conclude that the optimized TuTraSt parameters for studying CH_4 diffusion in zeolites is that of DS5 (g_{res} = 0.2 Å, E_{cutoff} = 60 kJ/mol, and E_{step} = 0.1 kJ/

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

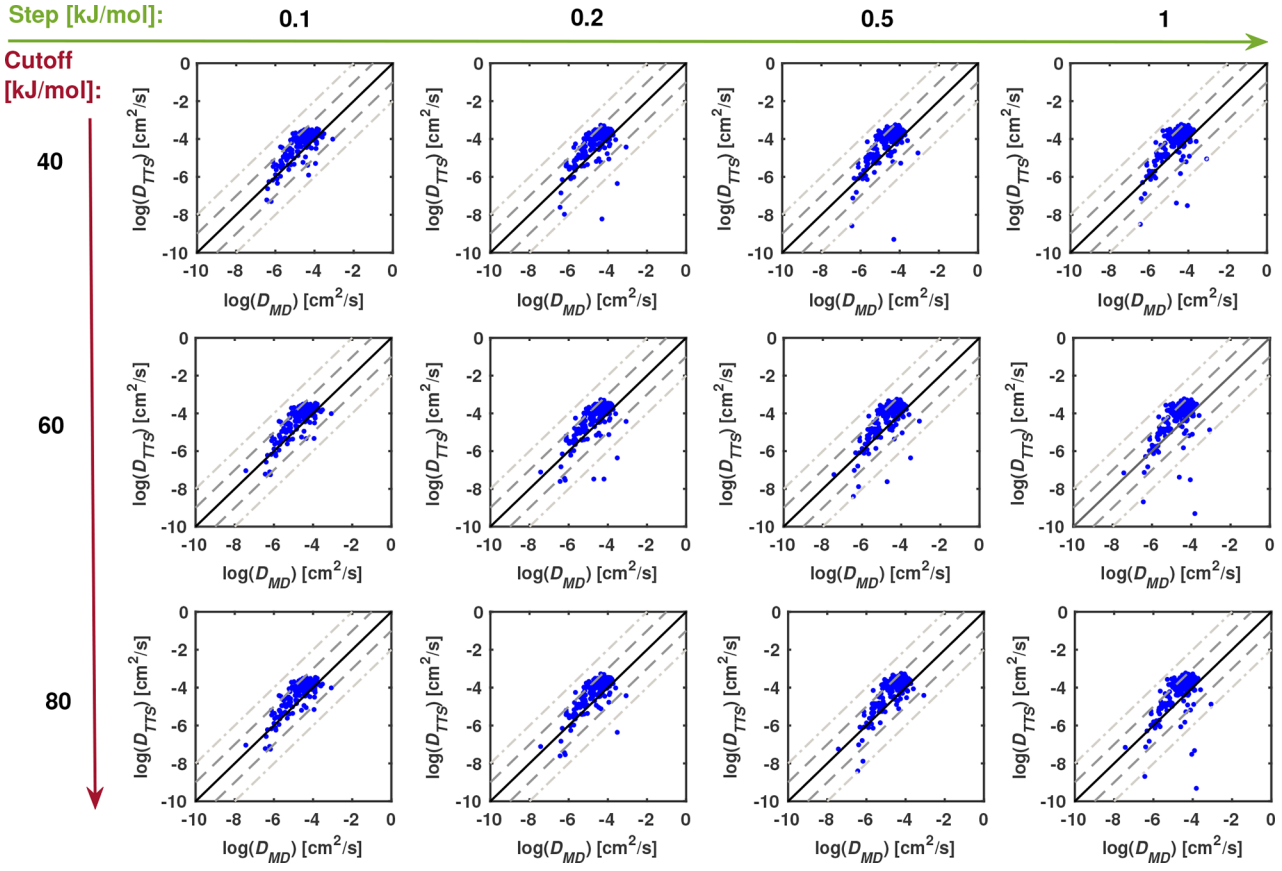


Figure 9. Comparing the D_{TTS} and D_{MD} for all IZA zeolites, using grids with grid size 0.2 Å. The plots differ by the choice of the values of E_{cutoff} and E_{step}.

Table 1. Metrics Describing Accuracy, Stability, and Speedup of the Diffusion Computations with the TuTraSt Algorithm Relative to MD for 6 Data Sets, All with E_{step} = 0.1 kJ/mol[a]

	g_{res}	E_{cutoff}	UF	nD_0	nD_{accept}	$nD_{outlier}$	ϵ	σ	$\rho_{Spearman}$	$\%_{accept}$	SU
DS1	0.1	40	42	134	149	14	0.44	0.42	0.73	83	585
DS2	0.1	60	45	134	148	12	0.42	0.42	0.72	83	409
DS3	0.1	80	54	130	146	12	0.41	0.44	0.71	81	363
DS4	0.2	40	0	139	179	21	0.33	0.65	0.67	94	8257
DS5	0.2	60	0	139	185	15	0.37	0.43	0.69	96	5345
DS6	0.2	80	3	136	185	15	0.43	0.44	0.69	95	4359

[a] g_{res} values are given in Å, and E_{cutoff} values are given in kJ/mol.

mol) where TuTraSt predicted the D_s values within 1 order of magnitude of the MD data for 96% of the 339 cell directions of the 113 zeolite structures and with an unchanged moderately high ranking correlation, $\rho_{Spearman}$ = 0.69. Of the remaining 4%, 50% were within 1.1 orders of magnitude, and the maximum deviation was 1.4 orders of magnitude. The mean time to compute D_s of CH_4 for one structure of this data set for all 3 dimensions using the TuTraSt algorithm is 93 min on a single CPU of an Intel Broadwell processor running at 2.6 GHz. We also note that the median compute time per structure is 28 min where the mean is strongly skewed by two significant outliers where the unit cells of these structures are exceptionally large (cell parameters >50 Å). The corresponding time averages for MD are 1.0×10^5 min.

The detection of tunnels, TS, and basin partitioning, as well as the organization of the TSs into components between basins together with the assignment of directions for the processes, heavily relies on neighborhood searches. It is therefore not surprising that we find a stronger influence on the computational cost of the number of points checked, mainly affected by the grid resolution and the maximal energy value, than on the number of levels that are stepped through. The cost of each level is dependent on the number of new grid points at the respective isovalue which decreases as the gradient of the energy function increases when higher energy states are reached, while the cost is polynomially dependent on the grid resolution. The number of grid points is thus the main dependency of the computation cost. However, for a given system an accurate cost prediction cannot be made *a priori* as it is essentially not directly dependent on the total number of grid points but the number of grid points within E_{cutoff} and the number of independent transitions within identified diffusion tunnels as computed by neighborhood searches in processes P2 and P6 shown in the flowchart. Processes P1, P3, P4, P5, and P7 have a negligible contribution to the total computational cost.

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

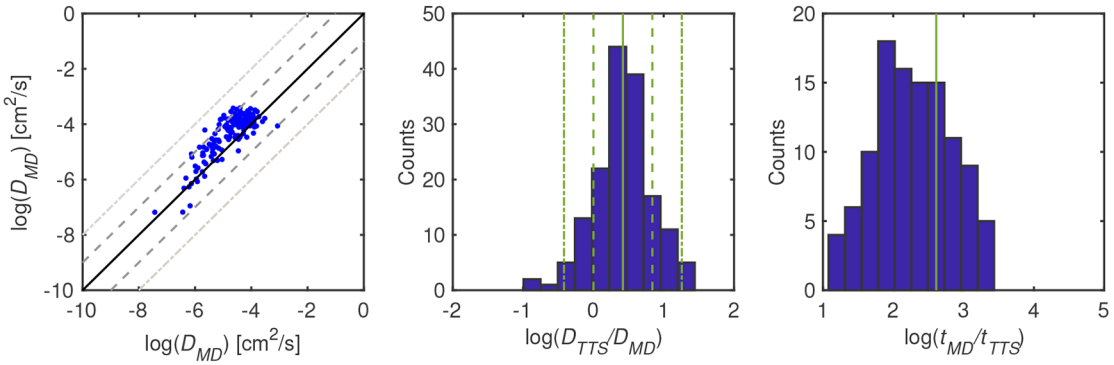


Figure 10. For DS2 the energetic parameters are g_{res} = 0.1 Å, E_{cutoff} = 60 kJ/mol, and E_{step} = 0.1 kJ/mol. The left panel shows the directional diffusion coefficients computed with the TuTraSt algorithm on the *y*-axis relative to the corresponding diffusion coefficients computed with MD on the *x*-axis on a log−log scale. The solid black line represents a perfect correspondence between the data sets, while the light gray and dark gray dashed lines guide the limits for deviation of 1 and 2 orders of magnitude, respectively. The center panel shows a histogram of the log-scale deviation from the MD validation set. Here the solid green line shows the ϵ = 0.42, and the dashed green lines show 1 (−) and 2 (.-) × σ = 0.42. The right panel shows a histogram of the log-scale speedup of a TuTraSt calculation per structure relative to the corresponding MD calculation. The green solid line shows the mean speedup of 409 for DS2.

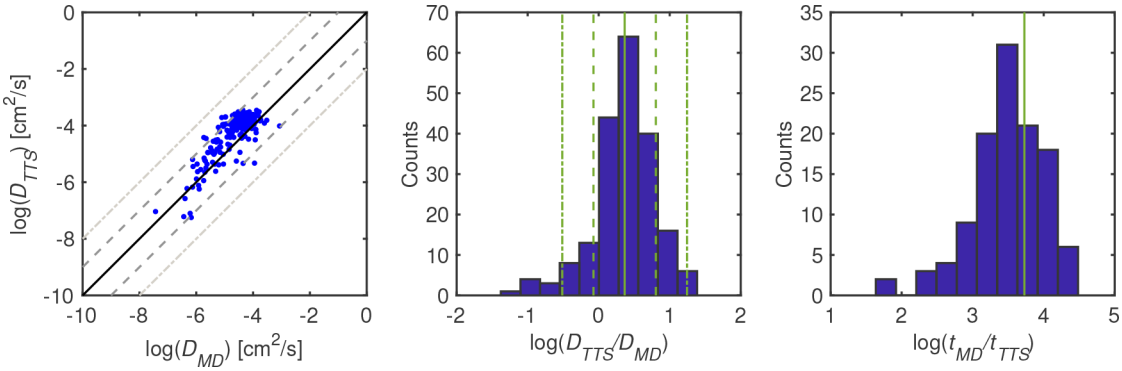


Figure 11. For DS5 the energetic parameters are g_{res} = 0.2 Å, E_{cutoff} = 60 kJ/mol, and E_{step} = 0.1 kJ/mol. The left panel shows the directional diffusion coefficients computed with the TuTraSt algorithm on the *y*-axis relative to the corresponding diffusion coefficients computed with MD on the *x*-axis on a log−log scale. The solid black line represents a perfect correspondence between the data sets, while the light gray and dark gray dashed lines guide the limits for deviation of 1 and 2 orders of magnitude, respectively. The center panel shows a histogram of the log-scale deviation from the MD validation set. Here the solid green line shows the ϵ = 0.37, and the dashed green lines show 1 (−) and 2 (.-) × σ = 0.43. The right panel shows a histogram of the log-scale speedup of a TuTraSt calculation per structure relative to the corresponding MD calculation. The green solid line shows the mean speedup of 5345 for DS5.

An additional note is that bias ϵ is consistently ∼0.4 indicating a systematic overestimation of the diffusion coefficients computed by the TuTraSt algorithm compared to the MD. This is expected considering we assume κ = 1/2 which is an ideal system where the TS are exact, meaning that recrossing events are not considered and hence our results will be an upper limit.

6. CONCLUSIONS, OTHER APPLICATIONS, AND OUTLOOK

We have presented a rigorous algorithm to analyze the geometry of three-dimensional potential energy landscapes felt by a diffusing particle in a solid-state material. The algorithm identifies tunnels and partitions the discretized energetic information up to a given energy cutoff value into basin and transition state "packages", respectively. These packages provide the information necessary to compute transition rates for all transitions between connected basins separated by transition state surfaces where the basins are reduced to discrete points in the unit cell volume. The tunnel and transition state search algorithm outputs a list of all possible transitions between lattice sites along with corresponding transition rates and transition vectors. This is the information required to compute the diffusion coefficient for a given transition with transition state theory, and we integrate a kinetic Monte Carlo simulation into the TuTraSt algorithm for automated computation of the directional diffusion coefficients. From the validation study of CH_4 diffusion in zeolites we conclude that TuTraSt is successful in predicting diffusion coefficients within the set accuracy goal of 1 order of magnitude for a large majority of the data (96%) provided optimal energetic parameters. These calculations are computationally cheap with a median computation time per structure ∼0.5 h on a single CPU which truly opens up for massive high-throughput screening studies of diffusion in solid materials of all energy landscape geometries. These run times are expected to be decreased by potentially 2 orders of magnitude by translating the code, currently written in Matlab, into a Python code with integrated C++ neighborhood search, also making it independent of licensed software.

This approach is general and can be applied to any host material where the structure fluctuations can be considered to have a negligible effect on the diffusion of the guest particle,

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

which can be modeled as a single particle. Such screening studies to be anticipated with this approach are $Li^+/Na^+/Mg^+$ diffusion in hypothetical solid-state electrolyte materials for ion batteries, extra-framework ion diffusion in nanoporous materials and glasses, or Kr/Xe separation in nanoporous materials. It is in principle possible to apply this approach to guest molecules essentially modeled by multiple particles, such as CO_2. However, in these cases constructing a potential energy grid is not as straightforward since at each grid point one needs to average over all possible degrees of rotational freedom, which may become a computational bottleneck.

The TuTraSt algorithm can also be applied as an analytical tool to interpret results and free energy surfaces retrieved from molecular dynamics or Monte Carlo simulations where the provided qualitative and quantitative information on transition states and basins facilitates the understanding of processes behind diffusional trends. One such study carried out by the authors acted to provide quantitative data to support the hypothesis that the presence of a few CH_4 molecules lowers the energy barriers of neighboring CH_4 in the direction along the channel of M-MOF-74 (M = Mg, Ni, Zn) and increasing diffusion coefficients as shown by NMR experiments and molecular dynamics simulations.

The applications of the TuTraSt algorithm are not restricted to the study of diffusion. Other properties also depend on the levels where clusters merge or the levels where tunnels appear. One example is the analysis of scalar functions derived from the electron densities as in the quantum theory of atoms in molecules for the description of atoms and bonds. Typically an isovalue is chosen for the visualization and analysis of the densities above the isovalue.[64,65] (De)localization of electrons and bond types can be distinguished by the different values where these electron density clusters merge.[66] So far this is done by a postanalysis by hand. Since our algorithm is designed to find the values of merging clusters, it can be applied to such an analysis in an automatized way. In a separate study by the authors, TuTraSt was applied to analyze the scalar function obtained from the electron density called "density overlap regions indicator" (DORI)[67] of a molecule in a molecular crystal.[68] Bond types are differentiated by taking a reasonable DORI value and integrating over the points in the regions that correspond to one bond with values greater or equal to the chosen value. Since the bond types are most distinguished at the values where the basins of a bond merge into one cluster, the merging points as determined from the TuTraSt algorithm give a good choice for the value up to which to integrate. This allows for performing a quantitative theoretical analysis that confirms the observed modifications of the covalent bonds in this molecule.[69,70]

An additional anticipated application of TuTraSt is to improve the currently used schemes to find inaccessible pockets of guest molecules in solids based on structure geometry, e.g. Zeo++,[58] by using instead an energy based scheme. Finding and blocking these spaces are important for Monte Carlo simulations that sample energetically particle insertions in porous materials to compute equilibrium uptake and thermodynamic quantities in different conditions.[71] As the TuTraSt algorithm is designed to identify energetically accessible space, it is straightforward to also use this information to identify the inaccessible space within a structure.

For studying the diffusion of sufficiently low concentrations of guest particles, such as in the CH_4 case study, it is a safe assumption that the potential energy landscape is representative for the free-energy landscape as these are equal at the infinite dilution limit. However, with increased guest particle loading, guest−guest interactions can play an increasingly significant role on the diffusion. For cases, such as CH_4 in porous materials, where particles may interact within a pore, particle collisions will be the dominant loading effect and are known to slow down the diffusion. While for ionic particles, the Coulomb forces constitute the dominating guest−guest interactions. These effects essentially perturb the transition state landscape, and our future focus is thus to further develop the TuTraSt algorithm to account for particle loading, in account to these two effects.

To correct for collisional loading effects we will apply and evaluate the dynamically corrected TST methodology such as proposed by Beerdsen and Dubbeldam et al.[48,49] Here, the recrossing parameter κ in eq 4 can be defined as a function of the loading. In the dynamically corrected TST formalism, the transmission coefficient κ corrects for trajectories which cross the transition state from basin A but fail to end up in basin B, compensating the loading induced shift of reaction coordinates.

To include charge interactions in our TST based lattice model we will test two ways to add the Coulombic ion−ion interactions, which will both be tested toward different ionic loadings and evaluated in terms of accuracy and computational effort. The first alternative is to update $P(q^*)$ in eq 5 "on-the-fly" for each possible process at each step in the kMC simulation where $E(q^*)$ and $E(q)$ are recomputed to take into account the Coulombic energies of surrounding ions. The second alternative is to apply a one time analytic correction to the lattice model where the Coulombic correction at each site is set in relation to the occupation probabilities at each surrounding lattice site.

For zero loading, it is possible to analytically obtain the diffusion coefficients for the limit of infinite dilution if all rates between any pair of sides are known following the method given by Braun and Sholl.[72] The diffusion coefficients are obtained from the expansion of the smallest eigenvalue of a matrix obtained from the spatial Fourier transform of the transfer matrix taking average thermal population into account, in the limit of the wave vector running to zero. This method has not been computationally implemented to our knowledge. If we assume that particles only move between neighboring sites, the input required for this analytical analysis can be obtained from a TuTraSt analysis. If implemented, this approach can therefore be used alternatively to the kMC simulation to calculate diffusion coefficients based on the TuTraSt analysis.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b01255.

Mean square displacements and statistical error analysis of diffusion coefficients for structures PSI, IFY, and AEI (PDF)

Slides showing growth of basins with tunnel and TS identification (clusters grown in order: left, right, top, bottom; tunnel detected in middle figure of bottom line (red); points belonging to TSs indicated with yellow boundaries) (PDF)

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

Movie of growth of CH_4 potential energy surface, up to 60 kJ/mol, in PSI zeolite framework and transition states colored by energy levels (blue = 1−10 kJ/mol, turquoise = 11−20 kJ/mol, green = 21−30 kJ/mol, yellow = 31−40 kJ/mol, orange = 41−50 kJ/mol, and red = 51−60) (AVI)

Movie of growth of CH_4 potential energy surface, up to 60 kJ/mol, in IFY zeolite framework and transition states colored by energy levels (blue = 1−10 kJ/mol, turquoise = 11−20 kJ/mol, green = 21−30 kJ/mol, yellow = 31−40 kJ/mol, orange = 41−50 kJ/mol, and red = 51−60) (AVI)

Movie of growth of CH_4 potential energy surface, up to 60 kJ/mol, in AEI zeolite framework and transition states colored by energy levels (blue = 1−10 kJ/mol, turquoise = 11−20 kJ/mol, green = 21−30 kJ/mol, yellow = 31−40 kJ/mol, orange = 41−50 kJ/mol, and red = 51−60) (AVI)

AUTHOR INFORMATION

Corresponding Author

*E-mail: berend.smit@epfl.ch.

ORCID

Amber Mace: 0000-0002-0323-0210

Senja Barthel: 0000-0002-9175-5067

Berend Smit: 0000-0003-4653-8562

Author Contributions

†A.M. and S.B. are joint first authors.

Funding

The research of S.B. and B.S. was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 666983, MaGic). Part of the research was supported by the NCCR MARVEL, funded by the Swiss National Science Foundation. A.M. thanks the Swedish Science Council (VR) for financing (project number 2015-06320). The calculations were enabled by the Swiss National Supercomputing Centre (CSCS), under project ID s761. We acknowledge PRACE for awarding access to SuperMUC at GCS@LRZ, Germany.

Notes

The authors declare no competing financial interest.

The data that the results of this paper are based on is available on the Material Cloud Archive (https://www.doi.org/10.24435/materialscloud:2019.0011/v1). There we provide structural data, potential energy grids, and MSD data and plots from the TuTraSt analysis. For download and use of the TuTraSt code please visit our GitHub repository (https://www.doi.org/10.5281/zenodo.2586985).

ACKNOWLEDGMENTS

The authors thank Peter G. Boyd for making the energy grids used as input for the TuTraSt algorithm in the case study.

REFERENCES

(1) Koros, W. J.; Zhang, C. Materials for next-generation molecularly selective synthetic membranes. *Nat. Mater.* **2017**, *16*, 289−297.

(2) Sholl, D. S.; Lively, R. P. Seven chemical separations to change the world. *Nature* **2016**, *532*, 435−437.

(3) Kim, J.; Abouelnasr, M.; Lin, L.-C.; Smit, B. Large-Scale Screening of Zeolite Structures for CO2Membrane Separations. *J. Am. Chem. Soc.* **2013**, *135*, 7545−7552 PMID: 23654217 .

(4) Daglar, H.; Keskin, S. Computational Screening of Metal-Organic Frameworks for Membrane-Based $CO_2/N_2/H_2O$ Separations: Best Materials for Flue Gas Separation. *J. Phys. Chem. C* **2018**, *122*, 17347−17357.

(5) Qiao, Z.; Xu, Q.; Jiang, J. High-throughput computational screening of metal-organic framework membranes for upgrading of natural gas. *J. Membr. Sci.* **2018**, *551*, 47−54.

(6) Keil, F. J.; Krishna, R.; Coppens, M.-O. Modeling of Diffusion in Zeolites. *Rev. Chem. Eng.* **2000**, *16*, 71−197.

(7) Mace, A.; Hedin, N.; Laaksonen, A. Role of Ion Mobility in Molecular Sieving of CO2 over N2 with Zeolite NaKA. *J. Phys. Chem. C* **2013**, *117*, 24259−24267.

(8) Mace, A.; Laasonen, K.; Laaksonen, A. Free energy barriers for CO2 and N2 in zeolite NaKA: an ab initio molecular dynamics approach. *Phys. Chem. Chem. Phys.* **2014**, *16*, 166−172.

(9) Mace, A.; Leetmaa, M.; Laaksonen, A. Temporal Coarse Graining of CO2 and N2 Diffusion in Zeolite NaKA: From the Quantum Scale to the Macroscopic. *J. Chem. Theory Comput.* **2015**, *11*, 4850−4860 PMID: 26574273 .

(10) Gómez-Álvarez, P.; Calero, S. Highly Selective Zeolite Topologies for Flue Gas Separation. *Chem. Eur. J.* **2016**, *22*, 18705−18708.

(11) Balakrishnan, P.; Ramesh, R.; Kumar, T. P. Safety mechanisms in lithium-ion batteries. *J. Power Sources* **2006**, *155*, 401−414.

(12) Kato, Y.; Hori, S.; Saito, T.; Suzuki, K.; Hirayama, M.; Mitsui, A.; Yonemura, M.; Iba, H.; Kanno, R. High-power all-solid-state batteries using sulfide superionic conductors. *Nat. Energy* **2016**, *1*, 16030.

(13) Bachman, J. C.; Muy, S.; Grimaud, A.; Chang, H.-H.; Pour, N.; Lux, S. F.; Paschos, O.; Maglia, F.; Lupart, S.; Lamp, P.; Giordano, L.; Shao-Horn, Y. Inorganic Solid-State Electrolytes for Lithium Batteries: Mechanisms and Properties Governing Ion Conduction. *Chem. Rev.* **2016**, *116*, 140−162 PMID: 26713396 .

(14) Boyd, P. G.; Lee, Y.; Smit, B. Computational development of the nanoporous materials genome. *Nat. Rev. Mater.* **2017**, *2*, 17037.

(15) Tabor, D. P.; Roch, L. M.; Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Montoya, J. H.; Dwaraknath, S.; Aykol, M.; Ortiz, C.; Tribukait, H.; Amador-Bedolla, C.; Brabec, C. J.; Maruyama, B.; Persson, K. A.; Aspuru-Guzik, A. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **2018**, *3*, 5−20.

(16) Pophale, R.; Cheeseman, P. A.; Deem, M. W. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12407−12412.

(17) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-scale screening of hypothetical metal−organic frameworks. *Nat. Chem.* **2012**, *4*, 83−89.

(18) Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M. In silico Design of Porous Polymer Networks: High-Throughput Screening for Methane Storage Materials. *J. Am. Chem. Soc.* **2014**, *136*, 5006−5022.

(19) Mercado, R.; Fu, R.-S.; Yakutovich, A. V.; Talirz, L.; Haranczyk, M.; Smit, B. In Silico Design of 2D and 3D Covalent Organic Frameworks for Methane Storage Applications. *Chem. Mater.* **2018**, *30*, 5069−5086.

(20) Jain, A.; Ping Ong, S.; Hautier, G.; Chen, W.; Davidson Richards, W.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.

(21) Merkys, A.; Mounet, N.; Cepellotti, A.; Marzari, N.; Gražulis, S.; Pizzi, G. A posteriori metadata from automated provenance tracking: integration of AiiDA and TCOD. *J. Cheminf.* **2017**, *9*, 55−56.

(22) Baerlocher, Ch.; McCusker, L. B. *Database of Zeolite Structures*. http://www.iza-structure.org/databases/ (accessed 2018-11-28).

(23) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database − an open-access collection of crystal structures. *J. Appl. Crystallogr.* **2009**, *42*, 726−729.

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

(24) Allen, F. H.; Glusker, J. P. Preface. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, Preface.
(25) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171−179.
(26) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal−Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26*, 6185−6192.
(27) Tong, M.; Lan, Y.; Qin, Z.; Zhong, C. Computation-Ready, Experimental Covalent Organic Framework for Methane Delivery: Screening and Material Design. *J. Phys. Chem. C* **2018**, *122*, 13009−13016.
(28) Lee, Y.; Barthel, S. D.; Dłotko, P.; Moosavi, S. M.; Hess, K.; Smit, B. High-Throughput Screening Approach for Nanoporous Materials Genome Using Topological Data Analysis: Application to Zeolites. *J. Chem. Theory Comput.* **2018**, *14*, 4427−4437 PMID: 29986145 .
(29) Martin, R. L.; Simon, C. M.; Medasani, B.; Britt, D. K.; Smit, B.; Haranczyk, M. In Silico Design of Three-Dimensional Porous Covalent Organic Frameworks via Known Synthesis Routes and Commercially Available Species. *J. Phys. Chem. C* **2014**, *118*, 23790−23802.
(30) Simon, C. M.; Kim, J.; Lin, L.-C.; Martin, R. L.; Haranczyk, M.; Smit, B. Optimizing nanoporous materials for gas storage. *Phys. Chem. Chem. Phys.* **2014**, *16*, 5499−5513.
(31) Simon, C. M.; Kim, J.; Gomez-Gualdron, D. A.; Camp, J. S.; Chung, Y. G.; Martin, R. L.; Mercado, R.; Deem, M. W.; Gunter, D.; Haranczyk, M.; Sholl, D. S.; Snurr, R. Q.; Smit, B. The materials genome in action: identifying the performance limits for methane storage. *Energy Environ. Sci.* **2015**, *8*, 1190−1199.
(32) Kim, J.; Maiti, A.; Lin, L.-C.; Stolaroff, J. K.; Smit, B.; Aines, R. D. New materials for methane capture from dilute and medium-concentration sources. *Nat. Commun.* **2013**, *4*, 1694.
(33) Gómez-Gualdrón, D. A.; Wilmer, C. E.; Farha, O. K.; Hupp, J. T.; Snurr, R. Q. Exploring the Limits of Methane Storage and Delivery in Nanoporous Materials. *J. Phys. Chem. C* **2014**, *118*, 6941−6951.
(34) Gómez-Gualdrón, D. A.; Simon, C. M.; Lassman, W.; Chen, D.; Martin, R. L.; Haranczyk, M.; Farha, O. K.; Smit, B.; Snurr, R. Q. Impact of the strength and spatial distribution of adsorption sites on methane deliverable capacity in nanoporous materials. *Chem. Eng. Sci.* **2017**, *159*, 18−30.
(35) Fu, J.; Tian, Y.; Wu, J. Seeking metal−organic frameworks for methane storage in natural gas vehicles. *Adsorption* **2015**, *21*, 499−507.
(36) Bao, Y.; Martin, R. L.; Haranczyk, M.; Deem, M. W. In silico prediction of MOFs with high deliverable capacity or internal surface area. *Phys. Chem. Chem. Phys.* **2015**, *17*, 11962−11973.
(37) Lin, L.-C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. A.; Jariwala, K.; Rycroft, C. H.; Bhown, A. S.; Deem, M. W.; Haranczyk, M.; Smit, B. In silico screening of carbon-capture materials. *Nat. Mater.* **2012**, *11*, 633−641.
(38) Chung, Y. G.; Gómez-Gualdrón, D. A.; Li, P.; Leperi, K. T.; Deria, P.; Zhang, H.; Vermeulen, N. A.; Stoddart, J. F.; You, F.; Hupp, J. T.; Farha, O. K.; Snurr, R. Q. In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm. *Sci. Adv.* **2016**, *2*, e1600909.
(39) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. Structure-property relationships of porous materials for carbon dioxide separation and capture. *Energy Environ. Sci.* **2012**, *5*, 9849−9856.
(40) Qiao, Z.; Zhang, K.; Jiang, J. silico screening of 4764 computation-ready, experimental metal?organic frameworks for CO2 separation. *J. Mater. Chem. A* **2016**, *4*, 2105−2114.
(41) Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. What Are the Best Materials To Separate a Xenon/Krypton Mixture? *Chem. Mater.* **2015**, *27*, 4459−4475.
(42) Banerjee, D.; Simon, C. M.; Plonka, A. M.; Motkuri, R. K.; Liu, J.; Chen, X.; Smit, B.; Parise, J. B.; Haranczyk, M.; Thallapally, P. K. Metal−organic framework with optimally selective xenon adsorption and separation. *Nat. Commun.* **2016**, *7*, 11831.
(43) Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L.; Snurr, R. Q. Thermodynamic analysis of Xe/Kr selectivity in over 137,000 hypothetical metal-organic frameworks. *Chem. Sci.* **2012**, *3*, 2217−2223.
(44) Sumer, Z.; Keskin, S. Molecular simulations of MOF adsorbents and membranes for noble gas separations. *Chem. Eng. Sci.* **2017**, *164*, 108−121.
(45) Kahle, L.; Marcolongo, A.; Marzari, N. Modeling lithium-ion solid-state electrolytes with a pinball model. *Phys. Rev. Materials* **2018**, *2*, 065405.
(46) Voter, A. F. Introduction to the Kintetic Monte Carlo Method. In *Radiation Effects in Solids*; Dordrecht: 2007; pp 1−23, DOI: 10.1007/978-1-4020-5295-8_1.
(47) Auerbach, S. M. Theory and simulation of jump dynamics, diffusion and phase equilibrium in nanopores. *Int. Rev. Phys. Chem.* **2000**, *19*, 155−198.
(48) Beerdsen, E.; Smit, B.; Dubbeldam, D. Molecular Simulation of Loading Dependent Slow Diffusion in Confined Systems. *Phys. Rev. Lett.* **2004**, *93*, 248301.
(49) Dubbeldam, D.; Smit, B.; Vlugt, T.; Beerdsen, E. Molecular simulation of loading-dependent diffusion in nanoporous materials using extended dynamically corrected transition state theory. *J. Chem. Phys.* **2005**, *122*, 224712.
(50) Jørgensen, M.; Grönbeck, H. MonteCoffee: A programmable kinetic Monte Carlo framework. *J. Chem. Phys.* **2018**, *149*, 114101.
(51) Chen, C.; Lu, Z.; Ciucci, F. Data mining of molecular dynamics data reveals Li diffusion characteristics in garnet Li7La3Zr2O12. *Sci. Rep.* **2017**, *7*, 40769.
(52) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press, Inc.: Orlando, FL, USA, 2001; DOI: 10.1016/B978-0-12-267351-1.X5000-7
(53) Sprik, M.; Ciccotti, G. Free energy from constrained molecular dynamics. *J. Chem. Phys.* **1998**, *109*, 7737−7744.
(54) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(55) Jónsson, H.; Mills, G.; Jacobsen, K. W. *Classical and Quantum Dynamics in Condensed Phase Simulations*; pp 385−404, DOI: 10.1142/9789812839664_0016.
(56) Bennett, C. H. Exact Defect Calculations in Model Substances. In *Diffusion in Solids: Recent Developments*; Orlando, FL, USA, 1975; DOI: 10.1016/B978-0-12-522660-8.50007-6.
(57) Chandler, D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **1978**, *68*, 2959−2970.
(58) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149*, 134−141.
(59) Zimmermann, N. E. R.; Jakobtorweihen, S.; Beerdsen, E.; Smit, B.; Keil, F. J. In-depth study of the influence of host-framework flexibility on the diffusion of small gas molecules in one-dimensional zeolitic pore systems. *J. Phys. Chem. C* **2007**, *111*, 17370.
(60) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(61) http://lammps.sandia.gov (accessed: 2018-12-03).
(62) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(63) Martin, M. G.; Siepmann, I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.
(64) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: 1990.

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141

(65) Lepetit, C.; Fau, P.; Fajerwerg, K.; Kahn, M. L.; Silvi, B. Topological analysis of the metal-metal bond: A tutorial review. *Coord. Chem. Rev.* **2017**, *345*, 150−181.
(66) Marx, D.; Savin, A. Topological Bifurcation Analysis: Electronic Structure of CH. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 2077−2080.
(67) de Silva, P.; Corminboeuf, C. Simultaneous Visualization of Covalent and Noncovalent Interactions Using Regions of Density Overlap. *J. Chem. Theory Comput.* **2014**, *10*, 3745−3756.
(68) Meyer, B.; Barthel, S.; Mace, A.; Vannay, L.; Guillot, B.; Smit, B.; Corminboeuf, C. DORI Reveals the Influence of Noncovalent Interactions on Covalent Bonding Patterns in Molecular Crystals Under Pressure. *J. Phys. Chem. Lett.* **2019**. DOI: 10.1021/acs.jpclett.9b00220.
(69) Casati, N.; Kleppe, A.; Jephcoat, A. P.; Macchi, P. Putting pressure on aromaticity along with in situ experimental electron density of a molecular crystal. *Nat. Commun.* **2016**, *7*, 10901.
(70) Casati, N.; Genoni, A.; Meyer, B.; Krawczuk, A.; Macchi, P. Exploring charge density analysis in crystals at high pressure: data collection, data analysis and advanced modelling. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2017**, *73*, 584−597.
(71) Gómez-Álvarez, P.; Ruiz-Salvador, A. R.; Hamad, S.; Calero, S. Importance of Blocking Inaccessible Voids on Modeling Zeolite Adsorption: Revisited. *J. Phys. Chem. C* **2017**, *121*, 4462−4470.
(72) Braun, O. M.; Sholl, C. A. Diffusion in generalized lattice-gas models. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 14870−14879.

DOI: 10.1021/acs.jctc.8b01255
J. Chem. Theory Comput. 2019, 15, 2127−2141